

OPEN

CHIEF EDITOR'S NOTE: This article is part of a series of continuing education activities in this Journal through which a total of 36 *AMA PRA Category 1 Credits*[™] can be earned in 2017. Instructions for how CME credits can be earned appear on the last page of the Table of Contents.

The Power and Pitfalls of Big Data Research in Obstetrics and Gynecology: A Consumer's Guide

Amie Goodin, PhD, MPP,* Chris Delcher, PhD, MS,† Chelsea Valenzuela, BA,‡
Xi Wang, MPH, BM,§ Yanmin Zhu, BA,§
Dikea Roussos-Ross, MD,¶ and Joshua D. Brown, PharmD, PhD||

*Postdoctoral Associate, Department of Pharmaceutical Outcomes & Policy, University of Florida College of Pharmacy;

†Assistant Professor and ‡Student, Department of Health Outcomes & Policy, University of Florida College of Medicine;

§Graduate Student, Department of Pharmaceutical Outcomes & Policy, University of Florida College of Pharmacy; ¶Assistant Professor, Departments of Obstetrics & Gynecology and Psychiatry, University of Florida College of Medicine; and ||Assistant Professor, Department of Pharmaceutical Outcomes & Policy, University of Florida College of Pharmacy, Gainesville, FL

Importance: Research in obstetrics and gynecology (OB/GYN) increasingly relies on “big data” and observational study designs. There is a gap in practitioner-relevant guides to interpret and critique such research.

Objective: This guide is an introduction to interpreting research using observational data and provides explanations and context for related terminology. In addition, it serves as a guide for critiquing OB/GYN studies that use observational data by outlining how to assess common pitfalls of experimental and observational study designs. Lastly, the piece provides a compendium of observational data resources commonly used within OB/GYN research.

Evidence Acquisition: Review of literature was conducted for the collection of definitions and examples of terminology related to observational data research. Data resources were collected via Web search and researcher recommendations. Next, each data resource was reviewed and analyzed for content and accessibility. Contents of data resources were organized into summary tables and matched to relevant literature examples.

Results: We identified 26 observational data resources frequently used in secondary analysis for OB/GYN research. Cost, accessibility considerations for software/hardware capabilities, and contents of each data resource varied substantially.

Conclusions and Relevance: Observational data sources can provide researchers with a variety of options in tackling their research questions related to OB/GYN practice, patient health outcomes, trends in utilization of medications/procedures, or prevalence estimates of disease states. Insurance claims data resources are useful for population-level prevalence estimates and utilization trends, whereas electronic health record-derived data and patient survey data may be more useful for exploring patient behaviors and trends in practice.

Target Audience: Obstetricians and gynecologists, family physicians.

Learning Objectives: After completing this activity, the learner should be better able to identify and define terminology used in observational data research; compare the features, strengths, and limitations of observational study designs and randomized controlled trials; distinguish between types of observational data (eg, insurance administrative claims, discharges, electronic health record databases, surveys, surveillance data) and weigh the strengths and limitations of research that uses each data type; interpret and critique OB/GYN research that uses observational data and secondary data analysis; and gain exposure and familiarity with a selection of observational data sets used to study topics relevant to obstetrical and gynecological practice and/or health outcomes.

All authors, faculty, and staff in a position to control the content of this CME activity and their spouses/life partners (if any) have disclosed that they have no financial relationships with, or financial interests in, any commercial organizations pertaining to this educational activity.

Correspondence requests to: Amie Goodin, PhD, MPP, University of Florida College of Pharmacy, 1225 Center Dr, Gainesville, FL 32610. E-mail: amie.goodin@ufl.edu.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.obgynsurvey.com).

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The provision, administration, and evaluation of contemporary health care services generate an enormous quantity of accessible data for research. Access to “big data” by clinicians and researchers increases every year, with tools such as electronic health records (EHRs) and open-access publication of research data and findings gaining rapid availability.¹

Big data in health care often refer to EHR databases, patient registries, and administrative claims, among others. These data sources are also used in a variety of observational research study designs and clinical specialty areas. Research using these data may be undertaken by medical students or residents, junior investigators, or experienced research teams with varying levels of experience and knowledge using these types of data and study design approaches. This “consumer guide” will prepare readers to navigate and interpret research using observational data while providing explanations and context for related terminology. A second objective of this guide is to help the reader, either a junior investigator or an experienced clinician, critique studies in obstetrics and gynecology (OB/GYN) that use observational data, which will be accomplished by outlining how to assess common pitfalls of quasi-experimental study designs that use observational data. The guide then demonstrates the application of this knowledge using a hypothetical case study of an OB/GYN research article using observational data and a quasi-experimental, or nonrandomized, design. Finally, a compendium of observational data resources commonly used within OB/GYN research is provided.

METHODS

Terminology and definitions relevant to observational data research were identified via literature and keyword searches for terms using PubMed/MEDLINE and Google Scholar. Examples of the relevant identified terminology were based on reviews of the literature in OB/GYN research-focused journals. Categories of observational data were defined and reported in summary tables, with examples of studies using each type of data.

Additional Web searches were conducted to identify data resources, and the identified government-sponsored data collection agencies were cross-referenced to find further data resources. Researcher recommendations were also used to gather observational data sources, along with review of current literature for observational studies or secondary data analysis in OB/GYN-focused journals. The contents of each data resource were summarized, and the cost and accessibility of each resource were verified with the data provider or government data collection agency. A flowchart was constructed to allow

the user to quickly interpret the study design for a quasi-experimental study design using observational data.

RESULTS

Review of the literature resulted in the identification of 22 frequently used terms relevant to interpreting observational data research (Table 1). These terms will be used throughout the guide.

Selecting the Appropriate Study Design

The reader should use this section to aid in distinguishing quasi-experimental designs using observational data from randomized controlled trials (RCTs) or other experimental designs. Experimental studies, such as RCTs, rely on an investigator-assigned exposure, whereas an observational study collects or reviews data from an extant phenomenon or occurrence; that is, the exposure is assigned “naturally” by clinical decision-making processes, policy changes, and so on (Fig. 1). Figure 2 provides the reader with a “checklist” for quickly interpreting observational data research and will be consulted in later discussion to aid in the critique of an example study abstract.

For obstetrician-gynecologists, primary data collection and the use of RCTs are more appropriate for answering research questions such as determining whether a newly developed treatment or medication has the intended effect on patients. This type of research question requires a demonstration of causality, which can be more strongly inferred from RCTs because of the use of randomization, and is most often referred as an evaluation of “efficacy,” or answering the question of “Can this work?” Randomization of the “exposure” (for this type of research question, the exposure is the treatment or medication) reduces or eliminates the possibility of selection bias when determining what groups in the study population receive the treatment and what groups are used for comparison (eg, no treatment, placebo, or alternative treatment). This process of randomization allows the investigator to rule out observed effects for reasons other than the treatment. Randomization is thus a crucial component in the investigation of new medications and medical devices, which is why the US Food and Drug Administration requires the use of RCT study designs during the approval process. Randomized controlled trials maximize internal validity, that is, lack selection bias, but have limited external validity due to strict inclusion/exclusion criteria that limit the most sick patients and patients at extremes of age (very young or very old) and operate in a clinical environment that is not comparable to everyday practice.⁴² As such, many research questions can be left unanswered

TABLE 1
Useful Definitions in Observational Research

Term	Term Definition(s)	Example OB/ GYN Study
Study Designs		
Quasi-experimental design	Quasi-experimental designs have been described as "...nonrandomized, pre-post intervention studies." ² However, this term may be more broadly interpreted as any study design that does not explicitly use randomization or investigator-assigned exposure. So-called "natural experiments" are quasi-experimental because conditions arise that make it possible to observe differences over time and/or across groups without direct investigator intervention (eg, a study that reports observed outcomes following a change in clinical guidelines that are enacted in geographically similar locations at different points in time would be a quasi-experimental "natural" experiment).	Rowe et al ³ (2014)
Secondary analysis	Any use of data that were originally collected for another purpose; formally, "...the reanalysis of either qualitative or quantitative data already collected in a previous study, by a different researcher normally wishing to address a new research question." ⁴	Withagen et al ⁵ (2012)
Cohort study: retrospective	This type of study design also "...use[s] data already collected for other purposes... The study period may be many years but the time to complete the study is only as long as it takes to collate and analyze the data." ⁶ The "cohort" is defined by the investigator as a group that has some common trait of interest or is exposed to the same intervention. The cohort is followed forward after this exposure within the confines of the collected data. The term "retrospective" implies that data have been collected, and there is no active follow-up.	Nygaard et al ⁷ (1997)
Cohort study: prospective	The prospective cohort study is similar to a retrospective cohort study, but the cohort is defined at the beginning of the study period, and the investigator follows the cohort forward in time. Formally, "[a] research study that follows over time groups of individuals who are alike in many ways but differ by a certain characteristic (eg, female nurses who smoke and those who do not smoke) and compares them for a particular outcome (such as lung cancer)." ⁸	Macleod et al ⁹ (2008)
Case-control study design	"A study that compares 2 groups of people: those with the disease or condition under study (cases) and a very similar group of people who do not have the disease or condition (controls)." ⁸	Jacobsson et al ¹⁰ (2008)
Nested case-control design	A nested case-control study is a special case of a case-control study, where "...cases of a disease that occur in a defined cohort are identified and, for each, a specified number of matched controls is selected from among those in the cohort who have not developed the disease by the time of disease occurrence in the case." ¹¹ The "nesting" here refers to the performance of a case-control study design within the context of an existing cohort of participants.	Belghiti et al ¹² (2011)
"Large N" observational study	A large N (or n) observational study is so called because of the inclusion of a large study population or sample size. "Sample size refers to the number of subjects in a study. If there is only 1 sample, then the sample size is designated with the letter 'N.' If there are samples of multiple populations, then each sample size is designated with the letter 'n.' When there are multiple population samples, then the total sample size of all samples combined is designated by the letter 'N.'" ¹³ While there is no official definition for what constitutes a "large" study population, observational studies may potentially contain thousands (or even millions) of patients or observations, which results in highly powered statistical analyses.	Scheller et al ¹⁴ (2017)
Cross-sectional	"A study conducted at a single point in time by, so to speak, taking a "slice" (a cross section) of a population at a particular time." ¹⁵ Cross-sectional study designs are incapable of assessing temporality within the unit of time the data are sampled, but can be used to assess trends over the time unit.	Becker et al ¹⁶ (2006)
Ecological study	A study design used to "...compare clusters of people... by assign[ing] one exposure level for each distinct group [which]... can provide a rough estimation of prevalence of disease within a population." ¹⁷	Brotherton et al ¹⁸ (2011)

(Continued on next page)

TABLE 1. (Continued)

Term	Term Definition(s)	Example OB/ GYN Study
Interpreting Observational Studies		
Power	Power is often used to describe the ability of the statistical test(s) to accurately report estimates or effect sizes within the study and is influenced by the size of the study population or number of observations. Formally, “[t]he probability of correctly rejecting a null hypothesis that is false is called the statistical power (or simply, power) of the test.” ¹⁹ Observational study designs that use large population or sample sizes are typically higher powered than RCTs that use small sample sizes.	Stulberg et al ²⁰ (2011)
Validation (instrument and/or measure)	Validation may describe either “...the process of verifying the correct application of research measurements” ²¹ or the check on whether a particular instrument can be correctly applied in certain populations (eg, an obstetrical risk score that is applied or “validated” for effectiveness in assessing risk for both high- and low-risk pregnancy populations).	Bowling et al ²² (2010)
External validity	External validity describes how effectively you can apply the study results to other populations or to the general population; formally, the “...extent to which results of trials provide a correct basis for generalization to other circumstances.” ²³	Smeenk et al ²⁴ (2001)
Internal validity	Internal validity describes how well the study design and instruments used in the study avoid attributing observations to some confounding and/or unrelated factor; formally, it is the “...extent to which systematic error (bias) is minimized...” ²³	Elamin et al ²⁵ (2003)
Time series	“A research design in which measurements of the same variables are taken at different points in time, often with a view to studying social trends.” ²⁶ It should be noted that in some fields (eg, health economics and health policy) this term is used interchangeably with “longitudinal repeated measures.”	Crofts et al ²⁷ (2016)
Survival analysis	Survival analysis can be described as “...a statistical investigation of time to an event.” ²⁸ In health care research, survival analysis is so named because it is often used to assess time to death (mortality) such as in the graphing of Kaplan-Meier curves, but it should be noted that any time-to-event analysis could be conducted using this family of statistical analysis methods. In other fields, survival analysis is called <i>event history analysis</i> (health policy, psychology/psychometrics) or <i>duration analysis</i> (health economics).	Hart et al ²⁹ (1999)
Longitudinal	This term describes “...studies in which any designated group of persons was followed or traced over a period of time.” ³⁰ Longitudinal (or longitudinal repeated measures) as a term is common in health and epidemiology research, but other fields (eg, health economics) will use the term <i>time series</i> to describe the same study design.	Hecher et al ³¹ (2001)
Bias: selection	Selection bias is a type of “...biased allocation to comparison groups.” ²³ Selection bias is problematic because it could cause the masking of true effect in a study due to some confounding factor. RCTs avoid selection bias by randomly allocating subjects into comparison groups, but this may not be possible in most observational study designs. Observational study designs may address problems of selection bias in a variety of ways, some of which include the use of sophisticated statistical tools such as propensity score analysis.	Blessed et al ³² (2001)
Bias: nonresponse	Nonresponse bias is a problem typically observed in survey research. It occurs “...when sampled elements from which data are gathered are different on the measured variables in nonnegligible ways from those that are sampled but from which data are not gathered.” ³³ In other words, those who participate in the study may be different than those who choose not to participate in the study, and this causes bias in the study results.	Morgan et al ³⁴ (2006)

(Continued on next page)

TABLE 1. (Continued)

Term	Term Definition(s)	Example OB/ GYN Study
Other Relevant Terminology		
Imputation	Imputation describes a series of "...statistical process[es] that statisticians, survey researchers, and other scientists use to replace data that are missing from a data set due to item nonresponse." ³⁵ A common example of imputation is the calculation of body mass index from last assessed patient height and weight when current body mass index is missing in the data. Imputation techniques can also include extrapolation or modeling of missing data.	Steinauer et al ³⁶ (2008)
Sample weighting	"In survey research, a sample in which weighting has been applied to different categories (such as ethnicity) to make the sample more representative of the population." ³⁷ Sample weighting tools are useful for improving the generalizability of study results and are also practical tools for limiting data collection to small population subsets, which keeps data collection costs low.	Hladky et al ³⁸ (2011)
Stratification	Stratification is another tool typically used in survey research during sampling. "Stratification is a well-known sampling tool built on the premise that like units in a population should be treated similarly. It is a statistical fact that grouping similar units, when sampling, can generally reduce the variance of the survey estimates obtained." ³⁹	Chung et al ⁴⁰ (2012)
Clustering	Clustering may refer to either the application of a group label to a "cluster" of subjects who have some trait in common during statistical analysis, or ... "[i]n broad terms, <i>clustering</i> , or <i>cluster analysis</i> , refers to the process of organizing objects into groups whose members are similar with respect to a similarity or distance criterion. As such, a cluster is a collection of similar objects that are distant from the objects of other clusters." ³⁹	Nielsen et al ⁴¹ (2007)

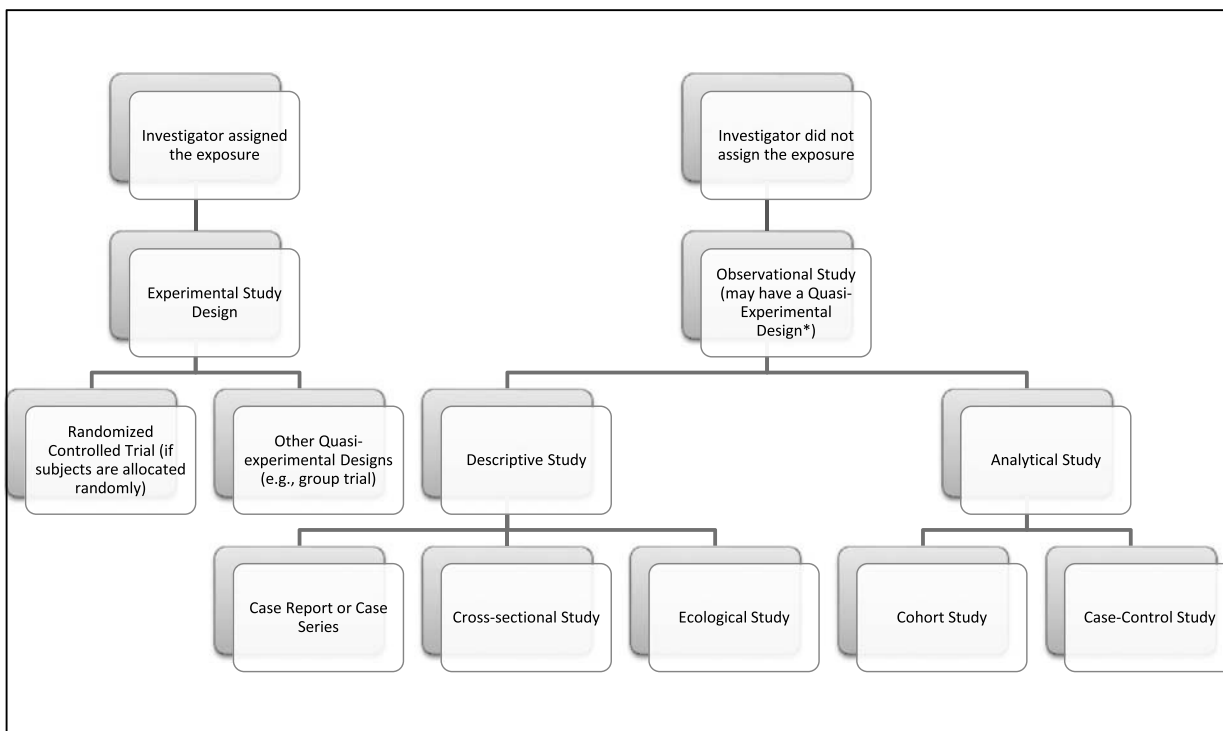


FIG. 1. Classifying study design. *If the investigator assigns the exposure, as in the case of a nonrandomized treatment versus control group study, then the study design would be classified as having an investigator-assigned quasi-experimental design. If the investigator does not assign the exposure in a treatment versus control group study, then the study design would be classified as an observational quasi-experimental design. An example of this is natural experiments, where some change like a new clinical guideline is enacted and the investigator compares the outcome on groups affected differently after the change.

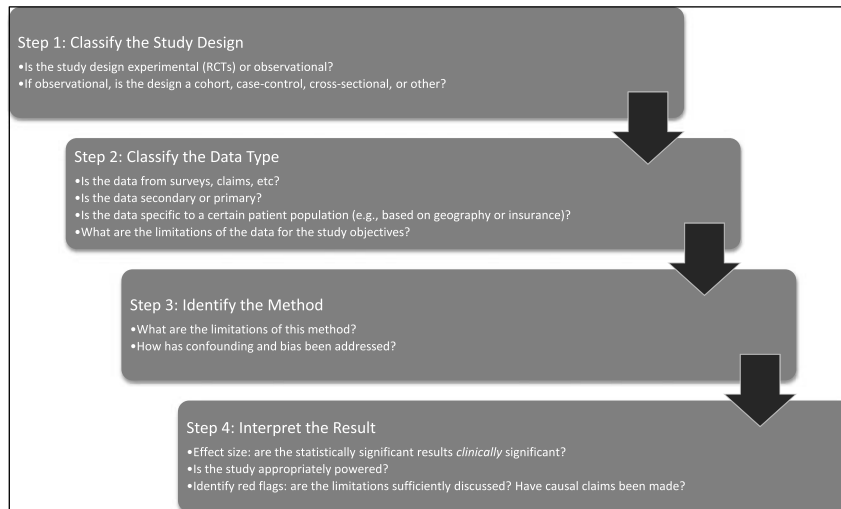


FIG. 2. A 4-step approach for identifying and interpreting observational studies.

in underrepresented or disparate patient groups and must be answered using “real-world” data and observational study designs.

Obstetrician-gynecologists also encounter research questions that may be preferable to investigate using observational study designs or secondary data analysis. Observational data are generated in the real world either through active data collection for research purposes or passive archiving of data for administrative purposes that can be reused. Real-world data are the product of everyday medical practice and not tightly controlled like that of an RCT. Patient medical treatments are part of a complex decision-making process and are non-randomized. Patients will receive treatments based on evaluation of risk factors, comorbid conditions, medical history, behavioral aspects, and other factors. Thus, there are significant threats to validity (ie, confounding and bias), associated with patient treatment and outcomes that have to be factored into study design and statistical analysis. However, observational data have strong external validity in the population from which they were generated and offer a relatively inexpensive and time-efficient means to conduct clinical research.

For example, it would be impractical, expensive, and potentially impossible to calculate the prevalence of cervical cancer in the general population by performing a cervical cancer screening and medical chart review for every female. We can, however, estimate cervical cancer prevalence using a variety of observational study design methods and data sources, such as querying patient registries, isolating claims for cervical cancer screenings and treatments in administrative data, surveying patients or clinicians or health systems, or conducting a secondary data analysis of reported surveillance data from government agencies. As another

example, an institutional review board would be hesitant to allow an RCT to compare birth defects caused by antidepressant use in pregnant women because of ethical and liability concerns. However, pregnant women often receive antidepressants in routine clinical practice and could be evaluated using observational data from a claims database linked to vital statistics. Observational data could then allow for a simultaneous comparison of all antidepressants, and active comparator treatment groups could be factored in to provide stronger evidence. Thus, confounding and bias concerns could be mitigated, for example, by strong study design and/or statistical adjustments using techniques such as regression, propensity score matching, or weighting. Such studies could be conducted in a timely manner using existing resources and without posing direct risk to patients.

Types of Observational Data

Seven types of observational data categories were identified from literature searches (Table 2), and each has several strengths and limitations. The summary definitions for these data types (surveys, admissions/discharge data sets, administrative claims data, registries, surveillance data, electronic medical records [EMRs], and linked data sets) are found in Table 2, with further discussion below. The strengths and limitations with the measurement of exposures, outcomes, and other explanatory factors are discussed in the following sections for each of the 7 types of observational data categories.

Surveys

Surveys constitute an entire family of investigatory tools that sample individual patients, clinicians, or health

TABLE 2
Types of Observational Data

Data Type	Description and Additional Terminology	Sample Data Set or Database	Example OB/GYN Citation
Survey	Surveys may include questionnaires or interviews of individual patients, clinicians, or health systems. Survey instruments and methodology are highly variable and are tailored to the target population and research question of interest.	NAMCS, NSDUH, NHIS	Laz et al ⁴³ (2012)
Discharges/admissions	Administrative data collected at admission or discharge to a health care facility. May include diagnoses, procedures, and patient demographic information. ⁴⁴	HCUP	Yasmeen et al ⁴⁵ (2006)
Administrative claims	"...[D]ata are collected for administrative or billing purposes, yet may be leveraged to study health care delivery, benefits, harms, and costs." ⁴⁶	Medicaid, commercial insurance claims	Biggs et al ⁴⁷ (2014)
Registries	"A registry is a collection of information about individuals, usually focused around a specific diagnosis or condition." ⁴⁴	US Zika Pregnancy Registry (US Centers for Disease Control and Prevention)	Muller and Miller ⁴⁸ (2017)
Population data (surveillance)	"Public health surveillance is the ongoing, systematic collection, analysis, interpretation, and dissemination of data about a health-related event for use in public health action to reduce morbidity and mortality and to improve health." ⁴⁹	PRAMS	Roberson and Hurwitz ⁵⁰ (2014)
EMRs	"...[A]n electronic record of health-related information on an individual that can be created, gathered, managed, and consulted by authorized clinicians and staff within one health care organization...." ⁵¹	Your local hospital EHR	Loudon et al ⁵² (2016)
Linked data sets	"Data linkage is the process of pairing observations from 2 or more files and identifying the pairs that belong to the same entity...." ⁵³		Salemi et al ⁵⁴ (2013)

HCUP indicates Healthcare Cost and Utilization Project; NAMCS, National Ambulatory Medical Care Survey; NHIS, National Health Interview Survey; NSDUH, National Survey on Drug Use and Health; PRAMS, Pregnancy Risk Assessment Monitoring System.

systems using questionnaires or interviews about a topic defined by the investigator. Sampling strategy, survey instrument design, respondent participation rates, and level of reporting subjectivity are all critical elements of interpreting observational study designs that use surveys. For these reasons, surveys tend to be one of the more complicated types of observational data to collect and interpret, and the resulting data are subject to varying levels of both validity and reliability. Secondary data analysis using publicly available survey data sources (eg, see National Ambulatory Medical Care Survey, National Survey on Drug Use and Health, and National Health Interview Survey in Supplemental Table 1, Supplemental Digital Content, <http://links.lww.com/OBGYNSURV/A28>) tends to be a better choice for OB/GYN investigators seeking to analyze survey data that were collected in a methodologically rigorous manner. These surveys are often weighted so that survey respondents, once the weights are applied in the statistical analysis, are representative of the general population. Survey data such as those collected by various government agencies are collected yearly

and also provide a useful source of data to track trends across time in the whole US population. These publicly available sources of survey data, however, come with limitations. None of the examples above were collected explicitly in populations relevant to OB/GYN research, although gender, sex, pregnancy status, and type/frequency of medical visits are available in each of these data resources. These surveys may not be appropriate for rare conditions or very restrictive patient populations because these may not have adequate sample sizes for study even with survey weights applied. In addition, survey responses may rely on unreliable self-reports and may not include data points relevant to your research question.

Administrative Claims Data

In recent years, large health care administrative claims databases have been widely accepted for research in OB/GYN. Administrative data are derived from automated electronic recording of encounters with the health care system, including physician's office

visits, hospitalizations, filled prescriptions, and diagnostic procedures, which are collected for administrative or billing purposes.^{46,55} These databases are readily available and relatively inexpensive to purchase/access and contain clinical information coded using accepted coding systems. They are also representative of routine clinical care and large populations.^{55,56} Claims data are usually generated from a single insured population (eg, Medicare, Medicaid, or private insurance), and the generalizability or external validity of research findings outside these discrete populations must be considered. Because administrative claims data are generated primarily for payment and administration of health services, these data are subject to inherent limitations because they are not collected for research purposes. For example, most administrative databases contain limited clinical information.⁵⁶ Physiological measurements such as blood pressure and glucose levels, tumor stage, and so on are generally not captured. Furthermore, diagnosis and procedure information may not be comprehensive enough and have low sensitivity and specificity for the condition that are meant to capture, and certain chronic conditions could be underdiagnosed. In addition, the reliability of information in administrative databases largely depends on its impact on payment and can differ between fee-for-service or capitated insurance benefits.⁴⁶ Treatment and services may not be reflected in the database if they are not covered by health insurance, such as elective cosmetic procedures or over-the-counter medications, or are simply paid in cash by the patient or administered for free (eg, vaccines or \$4 generic prescriptions). The practice of “up-coding” of patients in order to increase billing has been described and can be a major limitation to research, as well as constituting fraud.⁵⁷ When evaluating a study using administrative databases, it is imperative to recognize the strengths and the limitations of these data.

Despite its limitations, administrative claims databases have a wide spectrum of research applications. Using administrative claims data allows for the study of large patient populations, including pregnant women and infants/children, who tend to be partially or totally excluded from clinical trials.⁵⁵ Specifically, claims can be used to (1) examine drug utilization patterns (eg, the prevalence and duration of medication use during pregnancy), (2) study rare events (eg, breast cancers and congenital malformations), (3) evaluate appropriate prescribing (eg, antibiotic use and misuse among infants), and (4) link to other resources to expand the use of information in those databases (eg, linkage to birth certificate data and the National Death Index). Claims data are becoming more widely used to conduct comparative effectiveness and safety research, which

generally aims to compare alternative treatments where RCTs may never be performed or may be unfeasible. The US Food and Drug Administration's Sentinel Initiative is such an effort, developed as an active surveillance network built using a consortium of claims data providers. Claims data can be supplemented by linked data or be used to supplement RCTs or patient registries, which could provide substantial number data fields for further study or could extend the follow-up period beyond what was originally defined for a clinical study because of budgetary or other constraints.

Admissions/Discharge Data Sets

Health system admission or discharge data sets typically include diagnoses, procedures, and patient demographic characteristics from the hospital or health care facility in which the data were collected. These data are similar in format and content to administrative claims data and are thus subject to similar strengths and limitations. Discharge data are typically a 1-line, cross-sectional record of a hospitalization and are agnostic to when diagnoses and procedures occurred during the hospitalization. One important difference from administrative claims data is that admissions/discharges may include patients who have 1 or more types of insurance (or who have no insurance), which is typically not captured in administrative claims data that are derived from a single-payer source. Obstetrics and gynecology research that uses discharge data has found that some recorded procedures and diagnoses were highly accurate upon validation study for some conditions, but not as accurate for others.⁴⁵ The most commonly used discharge database, the National Inpatient Sample, is a random sample of hospitalizations that is weighted to be nationally representative but contains no unique patient-identifying identification variable. Thus, such results must be interpreted with caution as the unit of analysis is a “hospitalization” rather than a single patient, and some patients may be double counted. Such data are reliable for evaluating trends in utilization or costs and have been used to evaluate variation in care quality. Newer data resources, such as the National Readmissions Database, include patient identifiers and would allow one to follow individuals longitudinally in the data to evaluate 30-day readmissions, mortality, or other outcomes.

Registries

Registries are popular tools in OB/GYN research and are built to track a certain type of condition, use of medication/device, or patient within participating health systems. Registries are often useful for establishing

counts of observed adverse effects and outcomes and providing preliminary evidence for change over time in prevalence or incidence. One limitation is that registries tend to have voluntary enrollment, which means that participants might be inherently different from non-participants. Another is that registries are typically maintained by institutions with limited geographic penetration (eg, a single health system or professional organization).

Surveillance Data

Surveillance data are typically collected by a government or nonprofit agency to monitor a condition or attribute that is relevant to public health. Because a particular condition or attribute is the focus of data collection, an obstetrician-gynecologist may encounter limitations in using these data to answer research questions that require access to detailed clinical data. In addition, these data sources tend to be deidentified, which makes longitudinal tracking and data linkages difficult or impossible.

Electronic Medical Records

Many obstetrician-gynecologists have access to EMRs (or EHRs) used by their respective health care institutions. These data contain detailed clinical information and may also contain extended follow-ups and full medical histories. This depth of information is very valuable to OB/GYN research, but the use of these data for research purposes requires substantial review and oversight by your institutional review board and may require substantial time and staffing resources to filter and condense into usable data for analysis. In addition, the EMRs of a single health system are limited to the geographic area and population serviced by the health system, as well as the local policies and practices of the hospital and physicians, which limits the generalizability of study results to the population as a whole. Unless part of a fully integrated health system (eg, Kaiser Permanente or the Veterans Affairs system), patient visits to other physician offices or admissions to other hospitals may not be captured in the medical record. Development of a medical record that follows a patient, or is easily accessible from a cloud-based system, is an eagerly anticipated development for researchers looking to use medical records because this will further centralize the data for these purposes.

Linked Data Sets

Data linkages serve an important purpose in OB/GYN research because this is the primary tool for tracking females and progeny within single data sets and across multiple data sources. For example, observational

study designs that require the linkage of state vital statistics data (birth certificates and death certificates) are a critical tool for calculating maternal and neonate mortality and morbidity. Use of a common identifier is required to conduct a data linkage, which is why data linkages are difficult to conduct using publicly available data sources (these data are typically deidentified). However, some publicly available observational data resources may conduct a linkage by requiring the investigator to use the services of an honest-broker third party who conducts the linkage and then removes all identifying data. Such policies will vary significantly based on government or institutional policies.

Observational Data Resources

Supplemental Table 1 (Supplemental Digital Content, <http://links.lww.com/OBGYN SURV/A28>) contains a summary of observational data resources that are commonly used in OB/GYN research. This list is not meant to be exhaustive, and in some cases, a single prominent example was included for brevity (eg, commercial insurance claims, where a variety of vendors sell or allow access to different sets of administrative claims from commercial insurers).

DISCUSSION

Many researchers are engaged in translating big data resources into actionable benefit for health care quality and effectiveness. One study using the UK National Health Service administrative data demonstrated that observational study designs using “big data” could effectively be applied to what the authors called “small data” (eg, single-institution administrative data).⁵⁸ In the following sections, additional pitfalls and benefits of observational data research are discussed, and then a case study is examined so that the reader may apply the terminology and interpretation of study design to evaluating an OB/GYN study abstract.

Additional Pitfalls of Observational Data Research

There are several elements of observational research that we have yet to discuss. While the potential power of big data to disseminate findings or to investigate research questions with larger population sizes is apparent, many urge caution regarding the pitfalls of conducting such research using health care data. Namely, large-scale collection and “data mining” (ie, a series of statistical techniques for filtering large databases to isolate a characteristic or trend of interest sometimes in the absence of a research hypothesis) practices with health care data may pose risks related to causal inference and patient privacy, and/or

findings may be susceptible to misinterpretation if inappropriate or unfamiliar study designs or statistical techniques are applied.⁵⁹ For many, the latter is sometimes referred to as the “black box” and is made easy by modern statistical software. In this section, we examine some additional pitfalls of working with observational data, namely, interpreting the relevance of effect size (eg, clinical vs statistical significance), the technology constraints associated with big data research, and the use of data designed for purposes other than research.

Effect Size: Interpreting Clinical Versus Statistical Significance

It is important for OB/GYN professionals to know how to interpret and implement results from observational data in their clinical decision making. Observational studies rely on statistical analysis, which simply address the acceptance or rejection of statistical hypothesis. Statistically significant differences are determined by the generally accepted level of probability (the “*P* level”), and *P* values provide guidance with respect to random error due to sampling rather than a true change of outcome between study groups. Readers should consider several assumptions when interpreting statistical significances, such as sample size or distribution of the population. The power to detect a given effect size, and observe a statistically significant result, is generally a product of sample size and variation. With larger samples available in big data, the likelihood of statistical significance is high. This means that nonchance differences will be reported as “statistically significant” in big data studies, but those differences may be so small that the effect size, or practical difference, has little value for application to clinical practice. In other words, the observed statistical significance from an observational study alone is insufficient to apply in the clinical practice and may be susceptible to misuse by health care providers.

The results of an observational study can be statistically significant, but too minimal to be clinically important. However, there is no standard or defined understanding of clinically relevant changes. “Clinical significance” is generally defined as the smallest meaningful change in an observed effect, but this does not establish a standard effect size to be considered as worthwhile by the practitioner to enact a change in clinical care.^{60,61} There is a large degree of subjectivity in the judgment of clinical significance that could direct the course of patient care because of disparities in patients' characteristics, differences across health care fields, clinical experience, and goals of clinical care.

The aware reader should not focus on measures of significance presented in research articles but should evaluate the effect size, and its corresponding confidence intervals, to determine if a given effect size is enough to elicit a change in clinical practice. Furthermore, readers should expect both relative effect sizes and absolute effect sizes from research in order to fully appreciate the magnitude of the effect in question.

Let us examine a hypothetical example for interpreting effect size and statistical significance. A big data observational study finds a statistically significant effect of an intervention to improve patient preventive care. The intervention in our example is a change in insurance coverage, which was enacted by a large managed care organization to fully cover and encourage annual Papanicolaou (Pap) smear examinations. Administrative claims data were reviewed before and after the change in insurance coverage for all adult female patients in the large managed care organization with samples of several thousand patients receiving examinations in a 5-year period prior to the change in coverage and several thousand additional patients receiving examinations in the 5-year period following the change in insurance coverage. The study reported 2 major findings, namely, a *P* value that demonstrated a statistically significant difference in the proportion of patients in the managed care organization receiving examinations following the insurance change and an odds ratio with confidence intervals for the likelihood of patients receiving abnormal Pap test results following the coverage change.

These 2 statistical reports represent very different findings with varying applicability to clinical practice. First, the *P* value demonstrating a statistically significant difference in the proportion of patients who received Pap tests following the change does indicate that more patients received the tests, but does not indicate how many more patients (effect size) received those tests. This *P* value has limited clinical applicability because it demonstrates neither the size of the intervention effect nor the desirability of that effect (eg, more frequent Pap testing could potentially result in improved patient outcomes in high-risk patients, but what about low-risk patients? Is more frequent testing worthwhile for the whole population?). Second, let us suppose that the odds ratio in our example was reported as follows: the odds of an abnormal Pap test result following the insurance change was 1.03 (95% confidence interval, 1.01–1.05), which means that abnormal test results were found 0.03 times more often following the insurance change. These findings do give the reader information about the effect size of the intervention (ie, the odds of getting abnormal test results went up by 0.03 times, or 3%, following the insurance change);

however, this result still lacks clinical context. We know that a greater proportion of patients received annual testing following the insurance coverage change, but we do not know if the additional testing resulted in more false-positives (also known as type I errors) or a true capture of additional patients with abnormal test results. It is also unclear whether a 3% change in effect size is large enough to warrant change in practice—this would remain a clinical decision.

This example brings to light another consideration for working with big data and statistics, that of type I error, or false-positives. There are numerous methods for safeguarding against type I errors while working with exploratory statistical testing, so best practices would indicate that a statistician or methodologist be consulted to ensure that the selected study design and analysis methods are appropriate for the data available to the researcher.

Technology Constraints and Big Data

“Big data” is aptly named, in part, because many sources of observational big data are dramatically larger in sample size than what is typically used in experimental designs (RCTs). For example, a typical RCT conducted in OB/GYN research may enroll a study population of 200 female patients, whereas an extract of administrative claims from a large commercial or public insurer data set may contain millions of female patients who fit the study inclusion criteria with tens of millions of individual medical or pharmacy billing records. A data set of this size is not easily stored or analyzed using conventional computer hardware and software. A large data set (or database) may require the use of storage servers, querying software, sophisticated statistical analysis software packages, and the assistance of database administrators and/or statisticians to appropriately manage and analyze data of that magnitude. Access to the infrastructure required to work with big data may be cost or time prohibitive for an obstetrician-gynecologist practicing outside a research institution environment with established resources. Other data sources mentioned, such as surveys or discharge data, can be managed on local personal computers and are amenable to commonly used software packages that are not technically restrictive (eg, SPSS).

Additional Benefits of Observational Data Research

In this section, we discuss some additional benefits of working with observational data. First, the methods and procedures typically used in secondary data analysis tend to be highly reproducible. This is especially true when working with publicly available data and when

scholarly publications require thorough reporting of the research protocols used by observational study designs. As an obstetrician-gynecologist engaged in research, reproducibility is of tantamount importance when confirming or dismissing emerging evidence. Efforts by several independent groups and scientific societies have sought to standardize reporting of observational studies. Readers should be familiar with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines as a checklist for reporting observational study methods.⁶² Other resources are available for evaluating research using administrative claims data specifically.⁶³

Next, the use of extant data sources is a cost-effective strategy for engaging in research activities. A data source that is both publicly available and kept updated with contemporary entries (eg, annual surveillance data collected by government agencies) is of high value, but has a low cost of access for the investigator. In some cases, the purchase of extant data sources (eg, commercial insurance administrative claims) may be a more cost-effective means to answering research question than collecting primary data for an RCT, which tend to be expensive even when conducted in small populations. Observational data sometimes can allow for research questions to be evaluated that are deemed unethical or risky for approval of a clinical trial but have occurred in the real world (eg, birth defects associated with antipsychotic medications).⁶⁴

Some observational data resources also allow for the use of large population sizes, which are desirable for the high statistical power that reduces uncertainty and produces more precise estimates of observed phenomenon. A well-designed, highly powered “large N” observational study may produce evidence of greater value than an underpowered “small N” RCT in some cases, particularly among populations where it is difficult to conduct RCTs (eg, pregnant women). To reiterate, however, readers should remain aware of statistical versus clinical significance in studies with large sample sizes.

Case Study Application: Evaluating an Abstract That Uses Observational Data

Now that the reader is familiar with terminology, definitions, and evaluation strategies for interpreting observational data research, we will apply the strategies discussed throughout this guide by critically evaluating the following study abstract relevant to obstetrician-gynecologists and other clinicians who treat women and/or neonates. The abstract uses observational data to calculate the incidence of neonatal abstinence

syndrome and quantifies access to care differences for pregnant women seeking opioid abuse treatment:

Objective: Incidence of neonatal abstinence syndrome (NAS) is increasing because of the rise in opioid use. Rural states such as Kentucky have been disproportionately impacted by opioid abuse, and this study determines NAS burden nationally and in Kentucky while quantifying differences in access to care between Appalachian and non-Appalachian counties.

Methods: Neonatal abstinence syndrome rates were calculated using national (2013) and Kentucky (2008–2014) National Inpatient Sample discharge data. Births were identified using the *International Classification of Diseases, Ninth Revision* code 779.5 and live birth codes V30.x–V38.x. Counties were classified as rural, micropolitan, or metropolitan using census data. Proximity analysis was conducted via mapping from ZIP code centroid to nearest opioid treatment facility. Distance to treatment facilities was calculated and then compared using nonparametric testing for counties by rural and Appalachian status.

Results: Neonatal abstinence syndrome cases tripled from 2008 to 2014 in Kentucky counties, with a 2013 NAS rate more than double the national NAS rate. Rural and Appalachian counties experienced an NAS increase per 1000 births that was 2 to 2.5 times higher than urban/non-Appalachian counties, with a greater number of NAS births overall in Appalachian counties. All opioid treatment facility types were farther from rural patients than micropolitan/metropolitan patients ($P < 0.001$), as well as farther for Appalachians versus non-Appalachians ($P < 0.001$, all facility types).

Conclusions: Neonatal abstinence syndrome burden disparately affects rural and Appalachian Kentucky counties, whereas treatment options are disproportionately farther away for these residents. Policy efforts to increase NAS prevention and encourage opioid abuse treatment uptake in pregnant women should address rural and Appalachian disparities.

Note: This abstract is reprinted with permission from the *Journal of Rural Health* (publisher: John Wiley and Sons, Inc; license no. 4151490014858).⁶⁵

Using Figures 1 and 2 as a guide, we have enough information from the study abstract to quickly dissect and interpret the study design and findings. First, we can note that the researchers did not assign an exposure, and subjects were not randomized into the study, which means that the study design is nonexperimental (or quasi-experimental). In addition, the study is retrospective and contains 2 distinct research objectives: to calculate incidence of neonatal abstinence syndrome in births and to quantify access to care for opioid abuse among pregnant women in a particular state. The design used to address the 2 research objectives would be best described as a descriptive, cross-sectional design, because of the lack of comparison or control groups and lack of longitudinal follow-up data of mothers or neonates and because the data source lacks individual patient identifiers.

Next, the reader should note that administrative discharge data from the state of Kentucky and the National Inpatient Sample (see Supplemental Table 1, Supplemental Digital Content, <http://links.lww.com/OBGYNSURV/A28>, entry 12 for additional information about this data source from the Healthcare Cost Utilization Project family of data sets) were used in this article. National data are randomly sampled and would be nationally representative once these weights are applied, which can be used to generate estimates of neonatal abstinence syndrome for the entire country. However, these estimates cannot be calculated by state because of the way data are sampled. Kentucky data are complete records for the state, although Kentucky has been hit particularly hard by the opioid abuse epidemic, is rural, and has a distinct Appalachian disparate region and thus may not be comparable to other states. Noting the use of discharge data to define the patient population and to isolate diagnoses using billing codes is important, because diagnosis codes may not accurately capture all cases of pregnancy, substance abuse, and neonatal abstinence syndrome in these populations. Neonatal abstinence syndrome could be misdiagnosed, could be not caused by opioids (as was the focus of that research), and could vary in severity, which is not discernible from a single *International Classification of Diseases, Ninth Revision* diagnosis code. The aware reader should evaluate whether the authors have supported their choice of coding algorithms used either with prior research or with clear acknowledgement of the potential limitations of the research.

The abstract findings report both incidence rates and P values for differences in distance to opioid abuse care facilities. However, the abstract's method section provides limited information about the statistical analysis tools used to calculate either of these findings in the

study population, which could be cause for concern. Without these details, it is difficult to assess if the study is appropriately powered and whether the results are clinically meaningful (eg, how much higher would neonatal abstinence syndrome incidence in the study location [Kentucky] need to be compared to nationwide incidence to convince me it is a real problem?). Other results in this article show that women in rural areas would be on average 20 miles farther away from pregnancy-specific treatment centers than women in metropolitan areas, a statistically significant finding. While 20 miles may be a burdensome drive, is there evidence that this could affect treatment-seeking behaviors? Is it enough to enact policy changes from state or local government? In such situations, it is recommended to delve into the full-text article to find the missing information necessary to interpret the results, including references, which put the effect size in relative terms. If this information is unavailable in the full text, exercise significant caution in your interpretation and application of these study findings.

CONCLUSIONS

Observational data research and big data can provide clinicians and researchers with a variety of options for conducting and interpreting OB/GYN research, with applications ranging from assessing patient health outcomes, identifying trends in utilization of medications or procedures, or for calculating prevalence estimates of disease states. Different types of observational data have varied strengths and limitations. For example, administrative claims data sources are useful for population-level prevalence estimates and utilization trends, whereas EHR-derived data and patient survey data may be more useful for exploring patient behaviors and trends in practice. In comparison with RCTs, quasi-experimental designs using observational data may afford the study of larger populations in a more cost-effective manner.

ACKNOWLEDGMENTS

Publication of this article was funded in part by the University of Florida Open Access Publishing Fund.

REFERENCES

- Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;309:1351.
- Harris AD, McGregor JC, Perencevich EN, et al. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc*. 2006;13:16–23.
- Rowe H, Sperlich M, Cameron H, et al. A quasi-experimental outcomes analysis of a psychoeducation intervention for pregnant women with abuse-related posttraumatic stress. *J Obstet Gynecol Neonatal Nurs*. 2014;43:282–293.
- Payne G, Payne J. Secondary analysis. In: *Key Concepts in Social Research*. London, England: Sage Publications, Ltd; 2004. Available at: <http://methods.sagepub.com/book/key-concepts-in-social-research>. Accessed May 23, 2017.
- Withagen M, Milani A, de Leeuw J, et al. Development of de novo prolapse in untreated vaginal compartments after prolapse repair with and without mesh: a secondary analysis of a randomised controlled trial: prolapse repair and effect on untreated vaginal compartments. *BJOG Int J Obstet Gynaecol*. 2012;119:354–360.
- Mann CJ. Observational research methods—cohort studies, cross sectional studies, and case-control studies. *Afr J Emerg Med*. 2012;2:38–46.
- Nygaard IE, Rao SS, Dawson JD. Anal incontinence after anal sphincter disruption: a 30-year retrospective cohort study. *Obstet Gynecol*. 1997;89:896–901.
- National Cancer Institute. *NCI Dictionary of Cancer Terms*. 2017. Available at: <https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=286693>. Accessed May 25, 2017.
- Macleod M, Strachan B, Bahl R, et al. A prospective cohort study of maternal and neonatal morbidity in relation to use of episiotomy at operative vaginal delivery. *BJOG*. 2008;115:1688–1694.
- Jacobsson B, Ahlin K, Francis A, et al. Cerebral palsy and restricted growth status at birth: population-based case-control study. *BJOG*. 2008;115:1250–1255.
- Ernster VL. Nested case-control studies. *Prev Med*. 1994;23:587–590.
- Belghiti J, Kayem G, Dupont C, et al. Oxytocin during labour and risk of severe postpartum haemorrhage: a population-based, cohort-nested case-control study. *BMJ Open*. 2011;1:e000514.
- Sample size. In: *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage Publications, Inc; 2010. Available at: <http://methods.sagepub.com/reference/encyc-of-research-design/n396.xml>. Accessed May 25, 2017.
- Scheller NM, Pasternak B, Mølgaard-Nielsen D, et al. Quadrivalent HPV vaccination and the risk of adverse pregnancy outcomes. *N Engl J Med*. 2017;376:1223–1233.
- Cross-sectional study. In: *Dictionary of Statistics & Methodology*. Thousand Oaks, CA: Sage Publications, Inc; 2005. Available at: <http://methods.sagepub.com/reference/dictionary-of-statistics-methodology/n457.xml>. Accessed May 25, 2017.
- Becker JL, Milad MP, Klock SC. Burnout, depression, and career satisfaction: cross-sectional study of obstetrics and gynecology residents. *Am J Obstet Gynecol*. 2006;195:1444–1449.
- Thiese MS. Observational and interventional study design types; an overview. *Biochem Med (Zagreb)*. 2014;24:199–210.
- Brotherton JM, Fridman M, May CL, et al. Early effect of the HPV vaccination programme on cervical abnormalities in Victoria, Australia: an ecological study. *Lancet*. 2011;377:2085–2092.
- Statistical power. In: *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, Inc; 2008. Available at: <http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n550.xml>. Accessed June 20, 2017.
- Stulberg DB, Dude AM, Dahlquist I, et al. Abortion provision among practicing obstetrician-gynecologists. *Obstet Gynecol*. 2011;118:609–614.
- Ben Hardy Defines Validation*. London; 2017. Available at: <http://methods.sagepub.com/video/ben-hardy-defines-validation>. Accessed June 20, 2017.
- Bowling CB, Greer WJ, Bryant SA, et al. Testing and validation of a low-cost cystoscopy teaching model: a randomized controlled trial. *Obstet Gynecol*. 2010;116:85–91.
- Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ*. 2001;323:42–46.
- Smeenk JM, Verhaak CM, Eugster A, et al. The effect of anxiety and depression on the outcome of in-vitro fertilization. *Hum Reprod*. 2001;16:1420–1423.
- Elamin S, Langhoff-Roos J, Boedker B, et al. Classification of perinatal death in a developing country. *Int J Gynaecol Obstet*. 2003;80:327–333.

26. *Time Series Design*. London, England: Sage Publications, Ltd; 2006. Available at: <http://methods.sagepub.com/reference/the-sage-dictionary-of-social-research-methods/n208.xml>. Accessed May 25, 2017.
27. Crofts JF, Lenguerrand E, Bentham GL, et al. Prevention of brachial plexus injury—12 years of shoulder dystocia training: an interrupted time-series study. *BJOG*. 2016;123:111–118.
28. van den Hout A, Tom B. Survival analysis and the frailty model. In: *The Sage Handbook of Multilevel Modeling*. London, England: Sage Publications Ltd; 2013:541–558. Available at: <http://methods.sagepub.com/book/the-sage-handbook-of-multilevel-modeling/n30.xml>. Accessed May 25, 2017.
29. Hart R, Molnar BG, Magos A. Long term follow up of hysteroscopic myomectomy assessed by survival analysis. *Br J Obstet Gynaecol*. 1999;106:700–705.
30. Tooth L. Quality of reporting of observational longitudinal research. *Am J Epidemiol*. 2005;161:280–288.
31. Hecher K, Bilardo CM, Stigter RH, et al. Monitoring of fetuses with intrauterine growth restriction: a longitudinal study. *Ultrasound Obstet Gynecol*. 2001;18:564–570.
32. Blessed WB, Lacoste H, Welch RA. Obstetrician-gynecologists performing genetic amniocentesis may be misleading themselves and their patients. *Am J Obstet Gynecol*. 2001;184:1340–1342.
33. Nonresponse bias. In: *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, Inc; 2008. Available at: <http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n340.xml>. Accessed May 25, 2017.
34. Morgan MA, Hawks D, Zinberg S, et al. What obstetrician-gynecologists think of preconception care. *Matern Child Health J*. 2006;10:S59–S65.
35. Imputation. In: *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, Inc; 2008. Available at: <http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n216.xml>. Accessed May 25, 2017.
36. Steinauer J, Landy U, Filippone H, et al. Predictors of abortion provision among practicing obstetrician-gynecologists: a national survey. *Am J Obstet Gynecol*. 2008;198:39.e1–39.e6.
37. Weighted sample. In: *Dictionary of Statistics & Methodology*. Thousand Oaks, CA: Sage Publications, Inc; 2005. Available at: <http://methods.sagepub.com/reference/dictionary-of-statistics-methodology/n2084.xml>. Accessed June 20, 2017.
38. Hladky KJ, Allsworth JE, Madden T, et al. Women's knowledge about intrauterine contraception. *Obstet Gynecol*. 2011;117:48–54.
39. Post-stratification. In: *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, Inc; 2008. Available at: <http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n388.xml>. Accessed June 20, 2017.
40. Chung GS, Lawrence RE, Rasinski KA, et al. Obstetrician-gynecologists' beliefs about when pregnancy begins. *Am J Obstet Gynecol*. 2012;206:132.e1–132.e7.
41. Nielsen PE, Goldman MB, Mann S, et al. Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. *Obstet Gynecol*. 2007;109:48–55.
42. Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?” *Lancet*. 2005;365:82–93.
43. Laz TH, Rahman M, Berenson AB. An update on human papillomavirus vaccine uptake among 11–17 year old girls in the United States: National Health Interview Survey, 2010. *Vaccine*. 2012;30:3534–3540.
44. Agency for Healthcare Quality and Research. *The Value of Hospital Discharge Databases*. Bethesda, MD: AHQR; 2005:102. Available at: https://www.hcup-us.ahrq.gov/reports/final_report.pdf. Accessed August 8, 2017.
45. Yasmeen S, Romano PS, Schembri ME, et al. Accuracy of obstetric diagnoses and procedures in hospital discharge data. *Am J Obstet Gynecol*. 2006;194:992–1001.
46. Cadarette SM, Wong L. An introduction to health care administrative data. *Can J Hosp Pharm*. 2015;68:232–237.
47. Biggs MA, Harper CC, Malvin J, et al. Factors influencing the provision of long-acting reversible contraception in California. *Obstet Gynecol*. 2014;123:593–602.
48. Muller WJ, Miller ES. Preliminary results from the US Zika Pregnancy Registry: untangling risks for congenital anomalies. *JAMA*. 2017;317:35–36.
49. Buehler JW, Hopkins RS, Overhage JM, et al. *Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks: Recommendations From the CDC Working Group*. 2004. Available at: <https://www.cdc.gov/mmwr/pdf/rr/rr5305.pdf>. Accessed August 8, 2017.
50. Roberson EK, Hurwitz EL. Prescription drug use during and immediately before pregnancy in Hawai'i—findings from the Hawai'i Pregnancy Risk Assessment Monitoring System, 2009–2011. *Hawaii J Med Public Health*. 2014;73:382–386.
51. Agency for Healthcare Research and Quality. Health information technology archive: electronic medical record systems. Available at: <https://healthit.ahrq.gov/key-topics/electronic-medical-record-systems>. Accessed August 8, 2017.
52. Loudon H, Nentin F, Silverman ME. Using clinical decision support as a means of implementing a universal postpartum depression screening program. *Arch Womens Ment Health*. 2016;19:501–505.
53. Agency for Healthcare Research and Quality. *Linking Data for Health Services Research: A Framework and Guidance for Researchers*. 2013. Available at: <https://www.effectivehealthcare.ahrq.gov/ehc/products/494/1746/registries-linking-records-methods-draft-131107.pdf>. Accessed August 8, 2017.
54. Salemi JL, Tanner JP, Bailey M, et al. Creation and evaluation of a multi-layered maternal and child health database for comparative effectiveness research. *J Registry Manag*. 2013;40:14–28.
55. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58:323–337.
56. Romano PS, Hussey P, Ritley D. *Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives*. Rockville, MD: US Department of Health and Human Services, Agency for Healthcare Research and Quality; 2010.
57. Jesilow P. The effects of fraud on the evaluation of health care. *Health Care Anal*. 2005;13:239–245.
58. Mann RP, Mushtaq F, White AD, et al. The problem with big data: operating on smaller datasets to bridge the implementation gap. *Front Public Health*. 2016;4:248.
59. Gonzales JA, Lietman TM, Acharya NR. The draw(back)s of big data. *JAMA Ophthalmol*. 2017;135:422.
60. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415.
61. Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77:371–383.
62. University of Bern. Institute of Social and Preventive Medicine. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). STROBE statement. Available at: <https://www.strobe-statement.org/index.php?id=strobe-home>. Accessed August 8, 2017.
63. Mothral B, Brooks J, Clark MA, et al. A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. *Value Health*. 2003;6:90–97.
64. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2007;26:20–36.
65. Brown JD, Goodin AJ, Talbert JC. Rural and Appalachian disparities in neonatal abstinence syndrome incidence and access to opioid abuse treatment. *J Rural Health*. 2017. doi: 10.1111/jrh.12251.